

Informe resumen del proyecto “Using Neural Networks to Estimate Wave and Wind parameters from an HF radar”

Realizado, presentado y calificado en la Universidad de Sheffield. Sheffield, Junio 2008.

.....

Víctor Sawaqed Herrero

.....

Colmenarejo, 1 de Noviembre de 2010.

Contenido

SOBRE ÉSTE DOCUMENTO **3**

INFORME PRELIMINAR

INTRODUCCIÓN	4
DATOS Y METODOLOGÍAS	5
Los datos	5
Metodología	6
Resultados	6
CONCLUSIONES AL INFORME PRELIMINAR	8

INFORME FINAL

ACERCA DEL SISTEMA DE RADAR Y LA BOYA	9
El Espectro Doppler	9
La frecuencia de emisión varía a lo largo del día	10
La ambigüedad en la dirección de las olas	10
METODOLOGÍA DE TRABAJO	10
EXPERIMENTOS	12
Estimación de la altura de las olas	12
Estimación de la dirección máxima	12
Estimación del período medio	13
CONCLUSIONES Y TRABAJO FUTURO	14

Sobre éste documento

El objetivo de este documento es satisfacer un requerimiento de la Universidad Carlos III de Madrid para la convalidación de Proyectos de Fin de Carrera (PFC) realizados en estancias internacionales reguladas por convenio, por el cual se demanda un resumen en castellano del mismo. El Proyecto de Fin de Carrera al que se hace referencia fue realizado y calificado durante el curso 2008/2009 como parte de la carga docente cursada en el periodo de beca Erasmus en la Universidad de *Sheffield*, Reino Unido.

El tema inicial del PFC distaba sobre el que actualmente trata, esto fue debido a que por restricciones administrativas se tuvo que elegir PFC perteneciente a la carrera de Matemáticas Aplicadas el cual, por petición personal (e iniciativa propia) fue adaptado y convertido en lo que es ahora.

La estructura de la investigación realizada durante el PFC consiste en:

- Un estudio preliminar de las herramientas con las que se va a trabajar, éstas son las Redes de Neuronas, en concreto el Perceptrón Multicapa y el Espectro Doppler como resultado de las mediciones de los radares en la costa.
- Un análisis del citado Espectro Doppler y de los datos (y origen de ellos) que lo componen.
- Una estrategia metodológica que permitirá llevar a cabo los experimentos.
- Un diseño de cada experimento acordando variables como el tamaño del Perceptrón, rango de datos del radar en función del día o del estado real del mar en un momento particular, etc.
- Un análisis e interpretación de los datos obtenidos razonado.
- Una definición de próximos pasos y mejoras.

Los tutores que siguieron semanalmente el desarrollo del PFC en la Universidad de *Sheffield* pidieron expresamente que se explicara de una forma extensa y bien fundamentada la teoría básica de Redes Neuronales que se iba a utilizar, dado el escaso conocimiento del que disponían acerca de este tópico.

En este documento se hará un resumen breve de cada una de estas fases omitiendo las ilustraciones que no sean estrictamente necesarias y nociones básicas del ámbito informático así como todas las referencias en las que se basa el trabajo ya que se entiende que el fin de este resumen es servir de apoyo a la comprensión del lector.

Durante el desarrollo del PFC, se exigió la entrega de un informe (en formato *paper*) preliminar, donde se recogían los tres primeros puntos citados anteriormente, y un *paper* final donde se completaban las etapas y se justificaba aquello que lo necesitara aludiendo al *paper* preliminar.

Informe Preliminar

Introducción

Los datos que se van a analizar en este trabajo provienen de dos fuentes, una es una boya que flota en alta mar a 60 kilómetros de la costa, que registra los datos relativos a las olas como la altura o el período de crestas en un mar de fondo. La segunda fuente es un sistema de radar HF, que básicamente tiene información acerca de las corrientes (dirección, velocidad) y también sobre el viento, pero estos datos no son directamente registrados de este radar. Se utiliza el efecto Doppler para medir la velocidad de la ola y la velocidad de la corriente. La salida del radar es un conjunto de datos de potencia en función de la frecuencia retropropagada, con ellos se puede trazar el llamado espectro Doppler, y se analiza para describir las olas, etc

Hay varios métodos matemáticos para encontrar las relaciones entre el espectro Doppler y mediciones de ondas. Algunas de ellas implican la integración de una parte del espectro de potencia Doppler. Con ellos podemos llevar a cabo trabajos de minería de datos en el espectro y, por ejemplo tratar de averiguar cuánto será la altura o la dirección.

Lo que se pretende es incorporar otro método, la teoría de Redes de Neuronas, en el proyecto, vamos a probar si este nuevo método es más eficaz que los existentes.

Para entender de forma simple sobre cómo utilizar las redes neuronales, podemos decir que una red neuronal es una caja negra con un número de entradas y varias salidas. Por ejemplo, si ponemos la potencia (dB) de n puntos seleccionados de los espectros Doppler de uno o más radares como entrada y hacemos algunos ajustes en esta caja negra, vamos a ser capaces de utilizarla para predecir cuál será la altura de las olas. Digamos que esta red neuronal ha encontrado la función que relaciona los espectros energéticos con la realidad que ocurre en el mar.

La decisión de utilizar un perceptrón multicapa se basa en que es un aproximador universal que es capaz de encontrar relaciones no lineales entre las entradas y salidas.

Datos y metodologías

Se tratarán los dos conceptos por separado.

Los datos

Los datos que tenemos tienen que ser analizados para comprobar que se ajustan a las necesidades de las redes de neuronas.

Los datos con los que se está trabajando provienen de dos fuentes: Una de ellas es la boya que flota en alta mar y que registra cada media hora los datos “reales” del mar. Esta boya proporciona los parámetros de las olas, pero no proporciona información sobre el viento. La otra fuente es el radar HF que se encuentra en la orilla. Este radar proporciona el espectro de potencia de la retropropagación de la radiación que él mismo emite al mar. Esto puede ser usado para obtener datos como la altura de las olas. El alcance de medición comprende 30km – 200km fuera de costa.

Estamos utilizando los datos recogidos en el radar y la boya situada en el Mar Céltico. El sistema de la antena se encuentra en dos lugares, *Castlemartin* y *Nabor Point*. El Mar Céltico es una región limitada por el canal de Bristol en el Este y la costa de Gales del norte. La agencia del DEFRA (agencia ambiental del Reino Unido) eligió éste lugar como el sitio para un estudio de despliegue operativo para una red de monitorización de olas en el Reino Unido. El despliegue terminó en junio de 2005 y ofrece datos en tiempo casi real a los servidores de la Universidad de *Sheffield* y de la empresa *Seaview Sensing Ltd*.

Se necesitan dos arrays de radares en diferentes lugares (y emitiendo hacia el mismo punto) para obtener información acerca de la dirección de las olas.

En el espectro Doppler, se identifican una serie de regiones cuyo estudio revela la siguiente información:

Los picos de primer orden dan información sobre si la corriente va a hacia el radar o se aleja de él. Si la frecuencia que tenemos es menor que la frecuencia de referencia (0,0) significa que el objetivo se aleja del radar. Lo contrario sucede si la frecuencia es mayor que la frecuencia de referencia del radar.

Por lo tanto las regiones de primer orden dan información sobre las corrientes y la dirección del viento. Las regiones de segundo orden dan información sobre las olas y la velocidad del viento.

Los datos se facilitan cada hora por el radar, resultando 1024 puntos a razón de 0.005Hz cada uno Hertzios por cada 15 kilómetros de distancia de la orilla. Los datos están en unidades lineales de potencia (Wattios) se deben convertir a la escala dB, ya que dB es logarítmica y atenúa en gran medida las diferencias entre el máximo y el mínimo para mejorar la representación gráfica.

Con respecto a la boya, los datos facilitados son los siguientes: Tiempo, Tz, W_Pdir, Tpeak, Hm0, W_SPR donde tiempo es la fecha y hora expresados en GMT. Tz es el tiempo medio entre las crestas (en segundos) y W_PDIR (en grados) es la direcci3n de los picos de las olas. Tpeak (en segundos) es el peŕodo. Hm0 (metros) es la altura de las olas significativa y, por ́ltimo, W_SPR (en grados) la extensi3n media.

Metodoloǵa

En ́ste punto, fue deseo de los tutores del proyecto el incluir conocimientos de base sobre redes de neuronas, que como se ha explicado en la secci3n “Sobre este documento” se omitir aqú.

Durante el proyecto se va a tratar de encontrar relaciones entre los datos de la boya y del radar usando redes neuronales para poder predecir el estado real del mar conociendo los datos que proporciona el radar valiéndose del efecto Doppler. Es imprescindible que los datos del espectro de los radares est3n en la misma fecha y hora que los datos de contraste de la boya registrados, y tambi3n, en el mismo punto de la costa en el cual ́sta flota.

En este informe preliminar, se analizarn los datos disponibles de una manera que se obtenga informaci3n sobre c3mo estos datos son aplicables y adecuados para su tratamiento con redes neuronales.

Resultados

Se utilizarn los datos proporcionados por la boya del d́a 1-Feb-2005 a las 00:00 lo que nos da:

- 01- Feb- 2005 00:00:00, 4.29, 320.62, 5.50, 1.39, 24.57
- Tz = 4.29s
- W_PDIR = 320.62 degrees
- Tpeak = 5.50s
- Hm0 = 1.39m
- W_SPR = 24.57 degrees

El primer resultado que obtenemos son las representaciones grficas de los espectros Doppler provenientes de los radares de *Castlemartin* y *Nabor Point* para ese instante particular. Fue imprescindible llevar a cabo la normalizaci3n y adaptaci3n de rango de dichos datos explicada en el apartado Los datos. Los resultados se muestran en las imgenes Imagen 1 e Imagen 2.

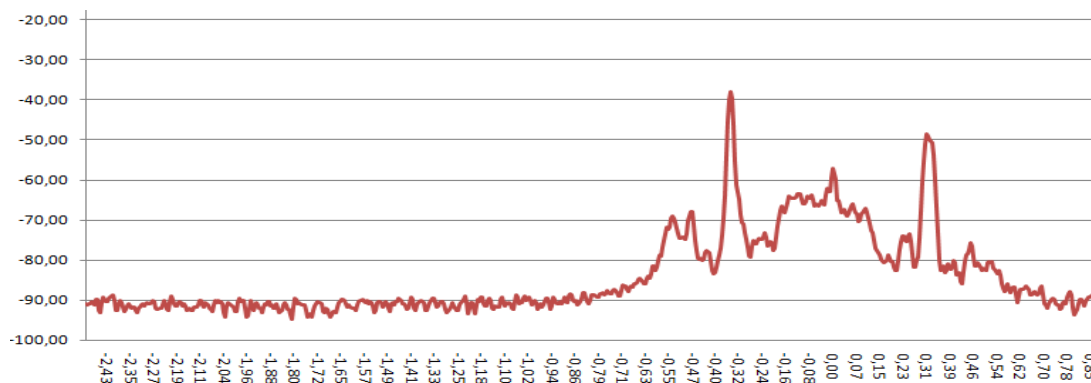


Imagen 1: Espectro Doppler a las 00:00 horas del día 01-Feb-05. Pertenece al radar de *Castlemartin* a 60.km. La frecuencia del radar era 10.363MHz.

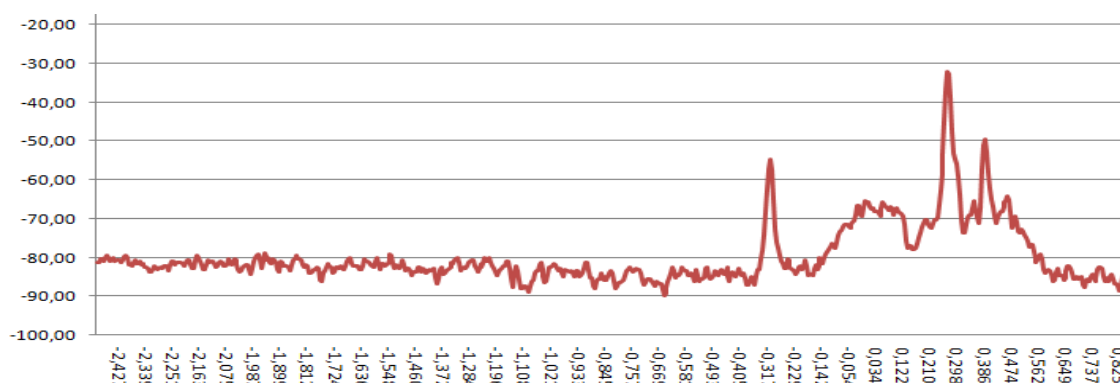


Imagen 2 Espectro Doppler a las 00:00 horas del día 01-Feb-05. Pertenece al radar de *Nabor Point* a 60.km. La frecuencia del radar era 08.036MHz.

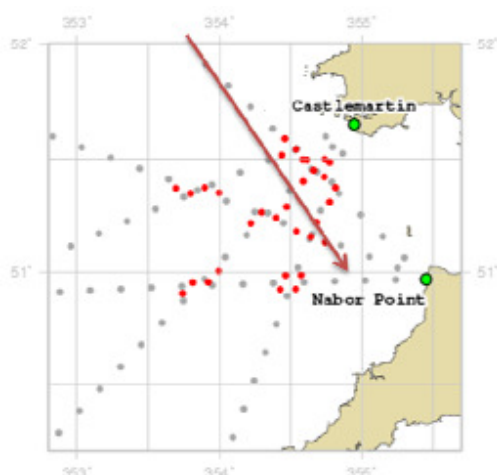


Imagen 3 Dirección aproximada de las olas según la Boya a las 00:00 del 1 - Feb - 2005

Como se observa en la Imagen 3 existe una relación entre los datos de la boya y de los radares. Mirando el espectro de potencia de *Nabor Point* en ese instante, vemos que el pico más alto se encuentra en una frecuencia positiva, por lo tanto sabemos que la dirección es **hacia** el radar. Si nos fijamos en espectro *Castlemartin* vemos más alto el pico en la frecuencia negativa por lo que se confirma que se **aleja** de este punto.

Sin embargo, si comparamos estos dos picos en las imágenes Imagen 1 e Imagen 2 vemos que la onda tiene una mayor componente en la dirección *Nabor*

Point que en la dirección de alejamiento de *Castlemartin*. Por lo tanto, podemos decir que la dirección debe ser como se muestra en la **¡Error! No e encuentra el origen de la referencia..** Concordando con lo medido por la boya.

Tambín puede extraer otra conclusi3n sobre la integridad de los datos: por el efecto Doppler sabemos que los dos picos en el espectro de enerǵa siguen esta f3rmula:

$$f_b(Hz) = \sqrt{\frac{4 \cdot g \cdot f \cdot \pi}{300}} \cdot \frac{1}{2 \cdot \pi}$$

Donde g es la aceleraci3n de la gravedad y f es la frecuencia del radar. Estos picos son siḿtricos respecto a cero (que es la frecuencia del radar). Por lo que uno es f_b y el otro es $-f_b$. Si se aplica esta f3rmula, por ejemplo, al espectro de *Nabor Point* resulta:

$$f_b = \sqrt{\frac{4 \cdot 9.8 \cdot 8 \cdot \pi}{300}} \cdot \frac{1}{2 \cdot \pi} = 0.29 \text{ Hz}$$

que es pŕcticamente el conjunto de puntos $[-0.3, 0.3]$, la ubicaci3n de los picos.

Y si aplicamos la f3rmula al espectro *Castlemartin*, obtenemos

$$f_b = \sqrt{\frac{4 \cdot 9.8 \cdot 10 \cdot \pi}{300}} \cdot \frac{1}{2 \cdot \pi} = 0.32 \text{ Hz}$$

Por ́ltimo, hay un pico que llama la atenci3n en el espectro de *Nabor Point* Imagen 2 entre las frecuencias y 0.386Hz 0.474Hz. Este pico puede existir por muchas razones, pero parece que podŕa ser un barco en movimiento hacia *Nabor Point*. Esta teoŕa est́ apoyada por el hecho de que en el espectro de potencia de *Castlemartin*, en 0 Hz de frecuencia podemos ver un peque1o pico, este podŕa ser el mismo barco en un movimiento totalmente paralelo a este radar por lo que no se produce el efecto Doppler pero si se recibe potencia de se1al retropropagada.

Conclusiones al Informe Preliminar

Los resultados muestran que podemos saber con seguridad que la selecci3n de los puntos precisos a 60km es totalmente relacionable con los datos proporcionados por la boya. Tambín encontramos que el procesamiento de los datos para ser tratados es correcto pues la formula concuerda (con un cierto error) con lo visualizado en el grafico.

Por tanto, recogen los requisitos para ser procesados por redes de neuronas: Ser consistentes, tener un set de datos para validar y ostentar una relaci3n.

La siguiente fase del trabajo consistiŕa en implementar una red neuronal y entrenarla con estos datos para ver c3mo se comporta la funci3n del error. Para realizar esto, tendremos que desarrollar peque1os programas que procesen la gran cantidad de datos de forma autoḿtica.

Informe Final

Acerca del sistema de radar y la boya

El efecto Doppler tiene lugar cuando las ondas de radio están siendo retropropagadas en una frecuencia diferente (mayor o menor) que la frecuencia de transmisión del radar. La cantidad de este incremento o decremento está relacionada con la velocidad de la ola del océano. Después, la señal recibida se procesa para producir el espectro de potencia. Este espectro se denomina espectro Doppler y es la fuente de datos para los experimentos en este proyecto.

En este proyecto las mediciones realizadas por la boya se consideran como los valores "reales".

En particular, los radares cuyos datos se van a usar, muestrean cada hora, lo que significa un diferente espectro Doppler por hora. La boya recoge los datos cada media hora.

El Espectro Doppler

Los espectros Doppler se crean con las mediciones del voltaje a través del tiempo. Puede verse un ejemplo en la Imagen 4.

Las medidas de potencia son devueltas por el radar en una escala lineal. Conviene trasladarla a una escala logarítmica (decibelios) para realizar un estudio gráfico de las regiones que se necesita (forma, el rango...). Eso se debe a que las frecuencias de interés tienen amplitudes mucho más pequeñas que los picos.

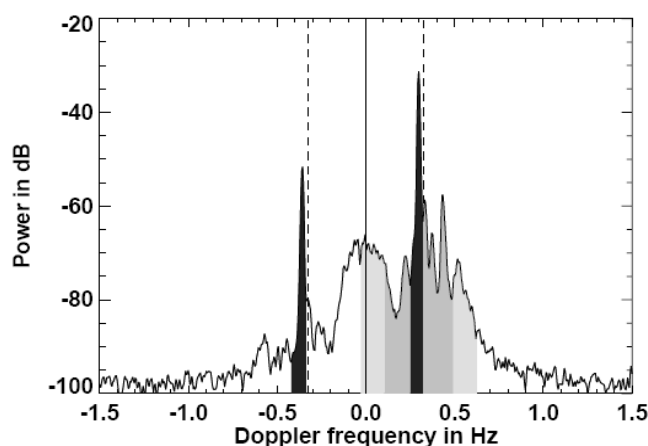


Imagen 4: Espectro Doppler. Las regiones sombreadas son las regiones de primer orden (black), las de segundo orden (dark grey) y otras regiones no utilizadas en este proyecto.

Los espectros Doppler tienen dos picos de primer orden que son mucho más altos que cualquier otro valor del conjunto de datos, ambos están simétricamente centrados en la frecuencia de referencia 0hz.

La frecuencia de emisi3n varía a lo largo del día

Frecuentemente para evitar interferencias, la frecuencia a la que emiten los radares varía a lo largo del día para evitar, de forma variable que el ruido interfiera en las mediciones. Debe ser tomado en cuenta que diferentes frecuencias harán que las regiones de segundo orden varíen su tamaño para una misma altura de olas. Por todo ello, incluir la frecuencia del radar en las mediciones resulta necesario.

La ambigüedad en la direcci3n de las olas

Para la misma altura de ola, la frecuencia retropropagada varía para el caso en que la ola esté viajando totalmente perpendicular a la direcci3n de emisi3n con respecto al caso en que viaje en la misma direcci3n. Esto provoca que la misma ola registre mediciones distintas en funci3n de su direcci3n.

Se necesitan dos algoritmos para estimar la altura de las olas, de tal modo que cada algoritmo procese el subconjunto de los datos correspondiente. Esto genera un problema de preprocesamiento de los datos. Se toma $\pm 10^\circ$ alrededor de la direcci3n exactamente perpendicular.

Metodología de trabajo

Para conseguir mejores resultados, tenemos que integrar las regiones. Para ello, se identifican cinco puntos en el espectro Doppler.

- f_b que es la frecuencia en la que el pico de primer orden está situado.
- f_l y f_r son las cotas de la regi3n de primer orden.
- f_{l2} y f_{r2} son las cotas de las regiones de segundo orden.

Los valores f_{l2} y f_{r2} , se obtienen de forma algorítmicamente diferente ya que están colocados a $\pm 60^\circ$ de la frecuencia de bragg.

Las integrales se sustituirán por sumas con la mayor precisi3n posible.

A parte de la integraci3n de las regiones de primer y segundo orden, otro dato que se extrae es el momento de las regiones de segundo orden. Para la regi3n izquierda $\frac{\sum_{f_{l2}}^{f_l} [D(f_i) \cdot (f_0 - f_i)]}{\sum_{f_{l2}}^{f_l} D(f_i)}$ y para la regi3n derecha $\frac{\sum_{f_r}^{f_{r2}} [D(f_i) \cdot (f_i - f_0)]}{\sum_{f_r}^{f_{r2}} D(f_i)}$.

Una vez se han identificado los datos que se van a utilizar en los experimentos y ćmo extraerlos, se procede a desarrollar de forma algortmica programas en Java que permitan extraer y formatear a fichero estos datos.

B́asicamente todos los programas que leen el conjunto de datos del espectro Doppler (que consiste en 1024 valores de potencia) de los archivos donde el radar vuelca la informacín (un archivo por hora), se almacena en un vector. Este vector se inspecciona para encontrar el ḿximo y, a continuacín, se lee desde esa posicín hacia la izquierda hasta que los valores dejan de decrementarse. Lo mismo se hace con el lado derecho. Aś se localizan los puntos fb y fl.

Otra subrutina integra recibiendo el vector con todos los valores y los ĺmites de la integracín (que podŕ ser fl - fr o fl2 - fr2). Con los archivos de la boya solamente hay que extraer los valores que se deseen si procesamiento intermedio. Después todo se une y formatea en un único archivo, cada ĺnea de este archivo es realmente un ejemplo de aprendizaje para la red neuronal.

Antes del entrenamiento los datos deben ser normalizados entre un rango que va de 0 a 1 en funcín de los ḿximos y ḿnimos de cada dimensín.

El SNNS (simulador de redes neuronales Stuttgart), se utiliza en este proyecto para llevar a cabo los experimentos. Este software permite crear muchos tipos de redes neuronales, aś como seguir el proceso de formacín. En particular, se utiliza la seccín para perceptrones multicapa MLP.

Por último hay dos pasos durante el experimento. El primero es seleccionar el ńmero óptimo de neuronas en la red neuronal, para esto se creaŕ una tabla comparativa en la que se tendŕn los diferentes errores de las diferentes redes de las cuales se elegiŕ la que de un menor error. La notacín de la configuracín de la red de neuronas en estas tablas estaŕ formada por ńmeros y guiones, por ejemplo, tres entradas, dos capas ocultas con diez neuronas y una salida seŕ 3-10-10-1.

El segundo paso es generar otra tabla que resume los resultados de la red elegida, esta tabla que contiene las estadísticas sobre los errores de prediccín.

Experimentos

Estimación de la altura de las olas

Análisis previos de los datos muestran empíricamente que a mayor altura de la ola, la energía registrada es mayor, con una correlación curvilínea.

Las redes neuronales deben tener tres entradas (dos para las regiones de segundo orden y otra para la frecuencia) y una salida (altura de las olas). En este experimento se utilizan las predicciones con los dos radares por separado.

Tras analizar varias configuraciones, se comprueba que las redes de Nabor Point tienen, en general, los resultados ligeramente mejor que las de Castlemartin. Se puede observar que durante el mes de febrero, la mayoría de las direcciones son entre 250 y 350 grados. Esta dirección es a través de Castlemartin (y por lo tanto, hacia / alejándose de Nabor Point), por lo que, por el principio de ambigüedad citado anteriormente, el radar de Nabor Point posee toda la información posible.

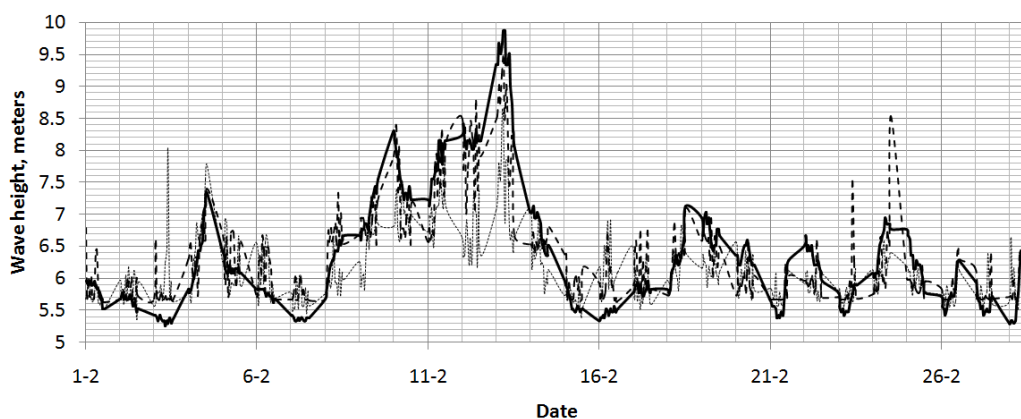


Imagen 5: Altura de las olas medida por la boya (línea de color negro sólido), la predicción del radar Castlemartin (línea gris punteada) y predicción de Nabor Point (línea discontinua negra) a lo largo de febrero.

En la Imagen 5 se puede observar una comparativa de la predicción hecha para este experimento.

Estimación de la dirección máxima

La dirección máxima es la dirección de las olas más altas.

De acuerdo con lo que se ha dicho acerca de la dirección y los espectros Doppler la sustracción de las regiones para integrar segundo del espectro en todo el pico más grande es el valor que se necesita para esta predicción.

La red neuronal ha de tener 4 entradas (la resta antes mencionados y la frecuencia de cada radar). La dirección máxima dada por la boya será la salida.

Tras una representación gráfica, Se observa que las direcciones que da la boyra se agrupan en dos grupos de ángulos. Dado que los grupos están en los extremos entonces, estos grupos podrán convertirse en uno si se tiene en cuenta que se trata de un círculo donde 0 y 360 corresponden a la misma dirección. Esta ambigüedad, podrá afectar el rendimiento de la red, ya que podrá confundir (dos espectros uno similar asociadas a 0 grados y el otro asociado a 360).

Con el fin de evitar esta ambigüedad, en primer lugar, se debe restar 180 a todos los datos para el rango de 0° a 360° se convierte en -180° a 180° y, se añade $+360^\circ$ a todos los valores entre -180 y 0 por lo que los valores en torno a 0° y 360 se encuentran cercanos.

A pesar de los errores de entrenamiento eran muy constantes. La red neuronal elegida fue la 4-8-1, debido al menor sesgo.

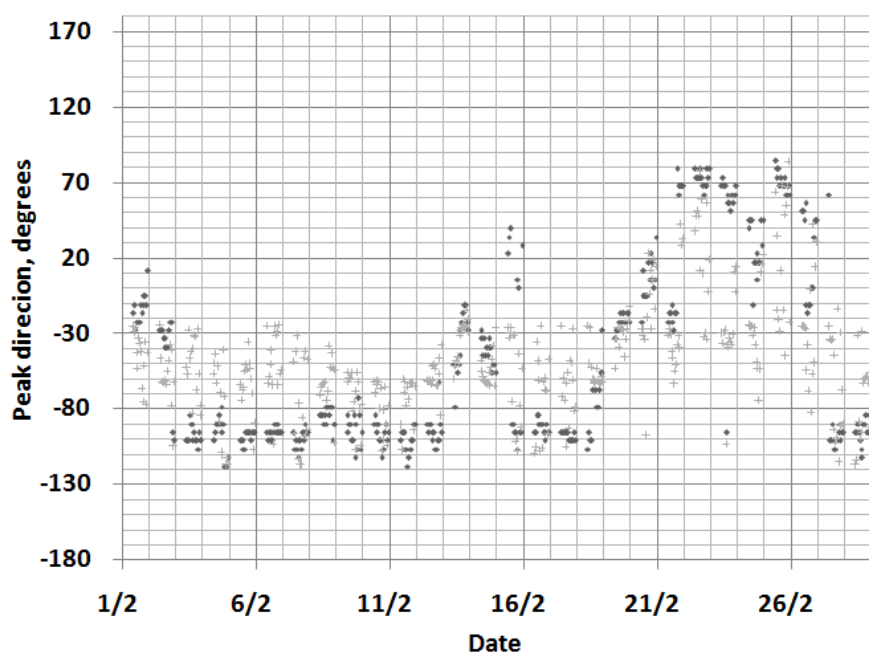


Imagen 6: Dirección máxima estimada por la boyra (negro sólido) y la predicción de la red neuronal (gris claro).

Estimación del período medio

El período medio es el tiempo en media que transcurre entre dos crestas de ola consecutivas. Existe una relación entre el primer momento de las regiones de segundo orden del espectro Doppler integrado a lo largo de las frecuencias del pico de primer orden.

Dado que los casos completamente perpendiculares son sólo un pequeño porcentaje del conjunto de datos de febrero, la ambigüedad no afectará mucho la predicción.

La red neuronal para este caso tendrá seis entradas y una salida. Las entradas son tres de cada radar, dos para los primeros momentos, uno de la región de segundo orden a la izquierda y uno de la derecha, y la frecuencia. Tras el proceso de aprendizaje, la configuración de red elegida fue 6-10-6-1 debido a la menor MSE y menor sesgo.

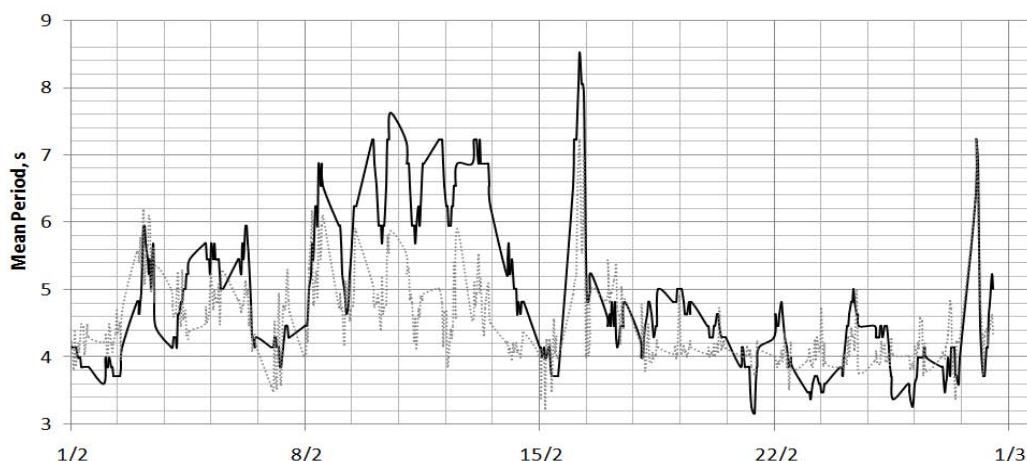


Imagen 7: Período medio medido por la boyas (línea de color negro sólido) y la estima con la red neuronal (línea gris de puntos) a lo largo de todo el mes de febrero de 2005.

Conclusiones y trabajo futuro

Los resultados de los experimentos muestran que las redes neuronales creadas no eran muy fiables. Un error medio de 1.09m en la predicción de altura de las olas con olas entre 5m y 10m en contra de 0.099m con la inversión dual o un error de 34° en la estimación de la dirección donde los algoritmos producen un error de entre 5 y 12 grados. Por último, un error de 0,2 s en la estimación del tiempo medio en contra de 0,69 segundos, cuando la duración varía de 4 a 8 segundos.

El método para elegir la configuración de las redes neuronales es más una manera empírica que los pasos de un método serio y teórico. Además, la red neuronal específica utilizada aquí podría ser un buen primer paso para el propósito de este proyecto, pero hay muchos más tipos de redes neuronales que podrían tomarse en cuenta.

Por lo tanto, como una sugerencia, la investigación debe ahondar mucho en el ámbito de las redes neuronales. Por ejemplo, podría centrarse en la predicción de altura de las olas y probar con varios tipos diferentes de redes neuronales o, probar otros métodos para la elección de la correcta configuración de la red.

Otra sugerencia podría ser el uso de una red neuronal para predecir en su conjunto los tres parámetros. La razón de esto es que los tres parámetros están usando las regiones integradas de segundo orden como fuente. Así, en esta red, las entradas serían dos (a la izquierda región de segundo orden y en la región derecha de segundo orden) y las salidas serían tres: la altura de ola, el período medio y la dirección máxima.